**Annals of Clinical and Analytical Medicine**

*Original Research*

# To what extent are the appendicitis scoring systems effective in preventing negative appendectomy. Comparison of nine scoring systems

Appendicitis scoring systems

Baris Mantoglu[1], Emre Gonullu[1], Guner Cakmak[1], Mertcan Akcay[1], Merve Yigit[1], Guven Erdogrul[2], Kayhan Ozdemir[3], Fatih Altıntoprak[4]
[1]Department of General Surgery, Sakarya University Training and Research Hospital, Sakarya
[2]Department of General Surgery, Mersin City Training and Research Hospital, Mersin
[3]Department of General Surgery, Urgup State Hospital, Urgup
[4]Department of General Surgery, Sakarya University Faculty of Medicine, Sakarya, Turkey

**Abstract**

Aim: Negative appendectomy is costly for both the patient and the health system; furthermore, it may introduce the patient to unnecessary surgical intervention and possible complications. This study aims to determine the most suitable one by comparing nine of the most popular appendicitis scoring systems.
Material and Methods: The study included 170 patients who were histopathologically diagnosed with appendicitis in the last year (Group 1), and 143 patients without appendicitis in the last five years (Group 2). The variables required to calculate scoring systems for the prediction of acute appendicitis were saved to the study datasheet, and each patient's score was calculated for each scoring system with the formulated excel file automatically.
Results: Among all scoring systems, the Karaman score was most efficacious at predicting appendicitis. The positive predictive value of the Karaman score was 89.9%, whereas the negative predictive value was 57.9%. The Alvarado score performed the best among the scoring systems. This was associated with a positive predictive value of 89.5%, negative predictive value of 85.7%, and sensitivity and specificity of 67.6% and 84.1%, respectively.
Discussion: The use of suitable scoring systems with or without imaging modalities or may reduce negative appendectomy rates.

**Keywords**

Appendicitis, Appendectomy, Negative Appendectomy, Score

DOI: 10.4328/ACAM.20932 Received: 2021-11-01 Accepted: 2021-12-06 Published Online: 2021-12-10 Printed: 2022-02-01 Ann Clin Anal Med 2022;13(2):210-214
Corresponding Author: Baris Mantoglu, Department of General Surgery, Sakarya Training and Research Hospital, Sakarya, Turkey.
E-mail: barismantoglu@gmail.com P: +90 505 815 93 79
Corresponding Author ORCID ID: https://orcid.org/0000-0002-2161-3629

## Introduction

Acute appendicitis (AA) is the most common diagnosis requiring emergency surgery worldwide, with a lifetime risk of 8.6% in males and 6.7% in females [1,2]. Despite developing laboratory and imaging methods, negative appendectomy rates vary between 3-25% in various publications [3–5]. The diagnosis of acute appendicitis is typically concluded by combining the patient's history, physical examination findings, and laboratory parameters within imaging modalities, when necessary. Unfortunately, a patient encountered in the emergency unit may set the surgeon in a tricky circumstance due to the lack of physical examination, and laboratory findings do not fully corresponds to acute appendicitis.

In challenging cases, the use of ultrasonography (USG) and computed tomography (CT) promotes the diagnosis. In the literature, the overall sensitivity and specificity of US are 76% and 95%, respectively, and 99% and 84% for CT, respectively. However, these imaging methods may not be simply accessible, as well as leading to high costs.

A negative appendectomy (NA) is costly for both the patient and the health system [6]; furthermore, it may introduce the patient to unnecessary surgical intervention and possible complications. To date, clinical scoring systems (CSS) have been defined to assist the diagnosis of AA [6–14]. These clinical scoring systems generally aim to clarify the diagnosis of AA based on scaled parameters of patients' examination findings and laboratory values. In the diagnosis of AA, the most precious of the CSSs should be the scoring system that reduces the rates of negative appendicitis as well as diagnosing positive appendicitis. Therefore, the study aims to determine the lowest rate of negative appendicitis among the nine popular scoring systems, as well as to reveal an effective scoring system that diagnoses positive appendicitis. Furthermore, the current study is the first in the literature to compare nine different CSSs.

## Material and Methods

The current retrospective comparative study included 170 patients whose pathology was confirmed as acute appendicitis in the last year at our institute, and 143 patients in whom the pathology did not reveal acute appendicitis from 1314 appendectomies in the previous five years. The ethics committee of our university approved the study with ethics number E-71522473-050.01.04-47683-410. Patients <18 years of age, pregnant patients, patients with existing malignancy, patients using steroids for any reason, immunosuppressive patients, COVID-19 positive patients, were excluded from the study. Each patient was examined by a consultant surgeon within the first hour after admitting to the emergency department. Following the laboratory analyzes of each patient, abdominal USG was performed on all patients. In the case of doubt about the diagnosis, an abdominal CT was performed. Lack of appendicitis in pathology reports was considered negative appendectomy. The histological diagnosis of appendicitis was set according to the infiltration of muscularis propria with neutrophils granulocytes. Variables required to calculate scoring systems (Alvarado, Raja Isteri Pengiran Anak Saleha Appendicitis (Ripasa), Tzanakis, Appendicitis inflammatory response (AIR), Eskelinen, Ohmann, Lintula, Fenyo-Lindberg, and Karaman scoring systems) for the prediction of acute appendicitis were saved to the study datasheet, and each patient's score was calculated for each scoring system with the formulated excel file automatically. The patients were divided into two groups as Group 1 positive appendicitis group and Group 2 as negative appendicitis group. Appendectomies were performed using conventional or laparoscopic methods. Statistical analysis was performed to compare the success of the scoring systems between groups.

### *Statistical analysis*

Descriptive analyzes were performed to provide information on the general characteristics of the study population. The Kolmogorov-Smirnov test was used to evaluate whether the distributions of numerical variables were normal. Accordingly, the independent sample t-test and the Kruskal Wallis test were used to compare the numeric variables between groups. Numerical variables were presented as mean ± standard deviation. Categorical variables were compared using the Chi-Square test. Categorical variables were presented as a count and percentage. A p-value <0.05 was considered significant. Receiver operator characteristic (ROC) curve analysis was used to identify the best cut-off value and assess the performance of the test score for appendicitis. Analyses were performed using SPSS statistical software (IBM SPSS Statistics, Version 23.0. Armonk, NY: IBM Corp.)

## Results

No difference was determined between the two groups included in the study in terms of age distribution. However, among laboratory parameters, white blood count (WBC), neutrophil percentage, PMNL rate were statistically high in Group 1 (Table 1). Total bilirubin value, on the other hand, was found to be high in group 2. Contrary to many studies, the C-reactive protein (CRP) value was high in group 1, but it did not reveal a statistically significant difference across group 2 (Table 1). Acute appendicitis was diagnosed in 82.9% of the patients who underwent USG. The consistency of male patients in group 1 and female patient distribution in group 2 is statistically significant.

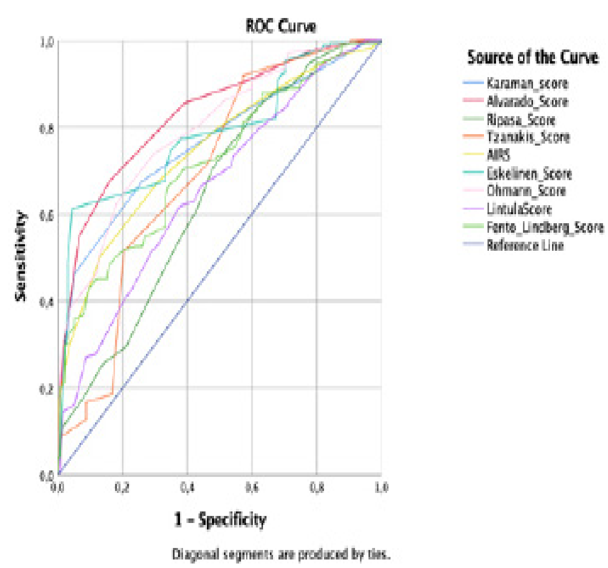


**Figure 1.** ROC curves for diagnostic performance of appendicitis scoring systems

Abdominal pain severity was evaluated high in Group 1 at a rate of 31.2% and assessed as mild in Group 2 at a rate of 43%. The increase in pain created a significant difference in Group 2 compared to Group 1. Besides, the duration of symptoms of 24-48 hours is statistically high in Group 2. Hyperactive bowel sounds, moderate abdominal rigidity, tenderness, and absence of rebound are statistically significantly less in Group 2.

When the nine clinical scoring systems included in the study were compared, the Ohmann score had the highest sensitivity with a rate of 74% (Table 2). When the area under the curve (AUC) was evaluated, the Alvarado score was determined first, and the Ohmann score was in the second line (Figure 1).

**Table 1.** Distribution of features related to appendicitis and not appendicitis groups.

| | | Appendicitis | Not Appendicitis | P value |
|---|---|---|---|---|
| | | 170 (54.3%) | 143 (45.7%) | |
| Age | | 33 (17-83) | 33 (18-79) | 0.60 |
| Neutrophil | | 10.45±4.58 | 5.57±2.82 | <0.05 |
| MPV | | 8.17±1.47 | 7.77±1.23 | 0.27 |
| Total Bilirubin | | 0.96 ± 0.64 | 1.26±8.26 | <0.05 |
| CRP | | 60.1±81.67 | 42.25±42.61 | 0.84 |
| WBC | | 14730.9±11449.84 | 11842±4491.67 | <0.05 |
| PMNL Ratio | | 75.91±13.9 | 63.64±12.67 | <0.05 |
| Negative Urine Analysis | No | 13 (18.3%) | 58 (81.7%) | <0.05 |
| | Yes | 157 (64.9%) | 85 (35.1%) | |
| App. In USG | No | 15 (12%) | 110 (88%) | <0.05 |
| | Yes | 155 (82.9%) | 33 (17.1%) | |
| Gender | Male | 110 (64.7%) | 54 (37.3%) | <0.05 |
| | Female | 60 (35.3%) | 89 (62.7%) | |
| Fever | No | 161 (56.1%) | 126 (43.9%) | 0.035 |
| | Yes | 9 (34.6%) | 17 (65.4%) | |
| Right Lower quadrant pain | Yes | 164 (96.5%) | 116 (81.0%) | <0.05 |
| | No | 6 (3.5%) | 27 (19.0%) | |
| The severity of pain | Mild | 30 (17.6%) | 62 (43.0%) | <0.05 |
| | Moderate | 87 (51.2%) | 69 (48.6%) | |
| | High | 53 (31.2%) | 12 (8.5%) | |
| Pain outside the right lower quadrant | No | 126 (74.1%) | 88 (61.3%) | 0.41 |
| | Yes | 44 (25.9%) | 55 (38.7%) | |
| Increasing pain | No | 76 (44.7%) | 30 (20.4%) | <0.05 |
| | Yes | 94 (55.3%) | 113 (79.6%) | |
| Pain transition umbilicus to RLQ | No | 80 (47.1%) | 69 (47.9%) | 0.57 |
| | Yes | 90 (52.9%) | 74 (52.1%) | |
| Vomiting | No | 109 (64.1%) | 89 (62.7%) | 0.72 |
| | Yes | 61 (35.9 %) | 54 (37.3%) | |
| Loss of appetite | No | 58 (34.1%) | 59 (40.8%) | 0.20 |
| | Yes | 112 (65.9%) | 84 (59.2%) | |
| Duration of symptoms | <24 h | 89 (52.4%) | 67 (47.2%) | <0.05 |
| | 24-48 h | 38 (22.4%) | 61 (42.3%) | |
| | >48 h | 43 (25.3%) | 15 (10.6%) | |
| RLQ pain with coughing | No | 56 (32.9%) | 34 (23.9%) | 0.17 |
| | Yes | 114 (67.1%) | 109 (76.1%) | |
| Bowel sounds with auscultation | None | 6 (3.5%) | 0 (0%) | <0.05 |
| | Normal | 161 (94.7%) | 126 (88.0%) | |
| | Hyperactive | 3 (1.8%) | 17 (12.0%) | |
| Rigidity | None | 35 (20.6%) | 46 (32.4%) | <0.05 |
| | Mild | 20 (1.8%) | 39 (26.8%) | |
| | Moderate | 74 (43.5%) | 43 (30.3%) | |
| | High | 41 (24.1%) | 15 (10.6%) | |
| Tenderness | No | 4 (2.4%) | 27 (19.9%) | <0.05 |
| | Yes | 166 (97.6%) | 116 (81.0%) | |
| Rebound | No | 45 (28.3%) | 85 (60.0%) | <0.05 |
| | Yes | 114 (71.7%) | 56 (40.0%) | |
| Rovsing | No | 101 (59.4%) | 76 (52.8%) | 0.34 |
| | Yes | 69 (40.6%) | 67 (47.2%) | |

MPV mean platelet volume, CRP C-reactive protein, PMNL polymorphonuclear leukocyte, App appendicitis, RLQ right lower quadrant

**Table 2.** Distribution of appendicitis diagnostic performance criteria of scoring systems

| | AUC | Cut-off | PPV | NPV | Specificity | Sensitivity | P value |
|---|---|---|---|---|---|---|---|
| Karaman Score | 0.765 | 4.5 | 0.899 | 0.579 | 0.739 | 0.676 | <0.05 |
| Alvarado Score | 0.831 | 5.5 | 0.895 | 0.857 | 0.841 | 0.676 | <0.05 |
| Ripasa Score | 0.645 | 91.75 | 0.642 | 0.947 | 0.572 | 0.607 | <0.05 |
| Tzanakis Score | 0.698 | 6.5 | 0.767 | 0.688 | 0.536 | 0.717 | <0.05 |
| AIRS | 0.749 | 5.5 | 0.898 | 0.800 | 0.688 | 0.676 | <0.05 |
| Eskelinen Score | 0.788 | 61.47 | 0.599 | 0.871 | 0.667 | 0.676 | <0.05 |
| Ohmann Score | 0.794 | 12.25 | 0.658 | 0.893 | 0.703 | 0.740 | <0.05 |
| Lintula Score | 0.659 | 16.5 | 0.684 | 0.583 | 0.623 | 0.618 | <0.05 |
| Fenyo Lindberg Score | 0.719 | 4.5 | 0.654 | 0.627 | 0.659 | 0.665 | <0.05 |

AUC: Area under curve, PPV: Positive predictive value, NPV: Negative predictive value

## Discussion

Despite advanced imaging methods, negative appendectomy rates have not decreased to the desired level in diagnosing appendicitis. Various examination and laboratory findings, including the female gender, are among the main reasons that make the diagnosis of appendicitis challenging. Unfortunately, many surgeons may face the dilemma of not performing unnecessary appendectomy or causing perforation due to delayed diagnosis in their professional practice. In the current study, our negative appendectomy rate was 10.88%, consistent with the literature.

The use of suitable scoring systems with or without imaging modalities or may reduce negative appendectomy rates. Furthermore, negative appendectomy can consequence in significant morbidity and mortality, such as prolonged hospital stay, postoperative surgical site infections, and even mortality [15]. On the other hand, diagnostic laparoscopy performed in suspicious cases also carries the risks of general anesthesia, and many surgeons eventually perform an appendectomy. The issue is that with or without inflammation of the appendix, postoperative complications are similar [16]. In this context, the described clinical scoring systems may effectively promote the diagnosis of appendicitis and reduce negative appendectomy rates in areas where imaging methods are difficult to access. Previous studies reveal that scoring systems reduce negative appendectomy rates [10,17–21]. The Alvarado score is the first of these scoring systems, and a score consisting of symptoms, signs, and laboratory parameters leads to diagnosis. In addition to similar parameters, different variables combined in the scoring systems defined later. Examples include gender in Fenyo-Lindberg, urinary symptoms in RIPASA and Ohmann scores, and the use of USG in the Tzanakis score. The Karaman score was the highest PPV value in the current study, and the RIPASA score had the highest NPV value. Moreover, Alvarado's score was the most specific, while Ohmann's score was the most sensitive (Figure 1). Based on AUC analysis, the Alvarado and Ohmann scores produced the highest AUC values among the nine scoring systems, while RIPASA and Lintula scores had the lowest AUC values (Figure 1). Considering the low valued AUC scoring systems, gender among the evaluation parameters draw attention. The inadequacy of this variable effects may be due to the presence or ratio of this parameter.

A significantly young female gender increases negative appendectomy rates due to numerous gynecological pathologies accompanying the differential diagnosis [22]. In addition, scoring systems are also affected by age and geographic population [23]. Unsurprisingly, in our study, the majority of group 2 consisted of the female gender.

Leukocytosis is the most critical laboratory value for the diagnosis of acute appendicitis so for scoring systems. In many studies, it has been stated that biochemical markers (C-Reactive Protein, bilirubin) may be helpful in the diagnosis of acute appendicitis. The fact that these markers differ in each study limits their routine use in diagnosing acute appendicitis; as expected, these values are not among the parameters in any scoring system. Our analysis has once again confirmed the influence of leukocytosis in acute appendicitis.

Although imaging modalities such as USG and CT are available to aid diagnosis and differential diagnosis, their utility is limited in accessibility, exposure to ionizing radiation, and cost-effectiveness. As evaluated in a previous study, negative appendectomy rates could not be reduced with the contribution of USG, but could be decreased at least by 8.3% with CT[24]. Precisely for this reason, it would not be off- target to state that the tomographic evaluation assists in the differential diagnosis further on the diagnosis of appendicitis. The USG was performed in all patients who were evaluated with a preliminary diagnosis of acute appendicitis. Patients still suspicious for appendicitis despite examination, laboratory data, and ultrasound had a CT scan. Despite this, unfortunately, our negative appendectomy rate was 10.88%.

From this point of view, combining CSS with ultrasound or tomography may reduce negative appendectomy rates. In the study conducted by Althoubaity et al., the higher rate of negative appendectomy detection (22%) using only USG in female patients showed that USG and scoring systems were more significant in females [24]. Likewise, in another recently published study, Ohmann, Lintula, and Ripasa reported that combining scoring systems with ultrasound could reduce the negative appendectomy rate up to 4% in female patients [25]. In addition, another study pointed out that, besides gender, age groups and geographic population also affect scoring systems in another study [23].

However, in hospitals where access to imaging methods is inconvenient or in cases where the cost cannot be condensed, CSS will significantly contribute to the diagnosis of appendicitis. In centers with no restrictions such as cost or accessibility in the usage of imaging methods, the combination of USG and/or CT with CSS may reduce negative appendectomy rates. In order to achieve the appropriate outcome, a CSS should be chosen as a result of studies carried out that minimize both gender, age, and other effects and can reduce negative appendectomy rates. In the current study, Alvarado and Ohmann score systems were the most effective scoring systems preventing negative appendectomy.

### *Study limitations*

Among the limitations of our study, in addition retrospective analysis, clinical scoring systems were compared not among patients who presented to the emergency department with abdominal pain but between groups who had appendectomy.

In conclusion, the use of appropriate CSS where access to imaging modalities is difficult, or in suitable centers addition to these modalities, may reduce negative appendectomy rates in the diagnosis of appendicitis. According to our analysis, Alvarado's and Ohmann score systems are the most relevant tests for this issue.

***Scientific Responsibility Statement***

*The authors declare that they are responsible for the article's scientific content including study design, data collection, analysis and interpretation, writing, some of the main line, or all of the preparation and scientific review of the contents and approval of the final version of the article.*

***Animal and human rights statement***

*All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. No animal or human studies were carried out by the authors for this article.*

***Funding:*** *None*

***Conflict of interest***

*None of the authors received any type of financial support that could be considered potential conflict of interest regarding the manuscript or its submission.*

***References***

*1. Chalazonitis AN, Tzovara I, Sammouti E, Ptohis N, Sotiropoulou E, Protoppapa E, et al. CT in appendicitis. Diagn Interv Radiol. 2008;14(1):19–25.*

*2. Jaschinski T, Mosch C, Eikermann M, Neugebauer EAM. Laparoscopic versus open appendectomy in patients with suspected appendicitis: a systematic review of meta-analyses of randomised controlled trials. BMC Gastroenterol. 2015; 15:48.*

*3. Mock K, Lu Y, Friedlander S, Kim DY, Lee SL. Misdiagnosing adult appendicitis: clinical, cost, and socioeconomic implications of negative appendectomy. Am J Surg. 2016; 212(6):1076–82.*

*4. Jeon BG. Predictive factors and outcomes of negative appendectomy. Am J Surg. 2017; 213(4):731–8.*

*5. Charfi S, Sellami A, Affes A, Yaïch K, Mzali R, Boudawara TS. Histopathological findings in appendectomy specimens: a study of 24,697 cases. Int J Colorectal Dis. 2014; 29(8):1009–12.*

*6. Khairy G. Acute appendicitis: is removal of a normal appendix still existing and can we reduce its rate? Saudi J Gastroenterol. 2009;15(3):167–70.*

*7. Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986;15(5):557–64.*

*8. Tzanakis NE, Efstathiou SP, Danulidis K, Rallis GE, Tsioulos DI, Chatzivasiliou A, et al. A new approach to accurate diagnosis of acute appendicitis. World J Surg. 2005; 29(9):1151–6.*

*9. Butt MQ, Chatha SS, Ghumman AQ, Farooq M. RIPASA score: a new diagnostic score for diagnosis of acute appendicitis. J Coll Physicians Surg Pak. 2014; 24(12):894–7.*

*10. Sitter H, Hoffmann S, Hassan I, Zielke A. Diagnostic score in appendicitis. Validation of a diagnostic score [Eskelinen score] in patients in whom acute appendicitis is suspected. Langenbecks Arch Surg. 2004; 389(3):213–8.*

*11. Lintula H, Kokki H, Pulkkinen J, Kettunen R, Gröhn O, Eskelinen M. Diagnostic score in acute appendicitis. Validation of a diagnostic score [Lintula score] for adults with suspected appendicitis. Langenbecks Arch Surg. 2010; 395(5):495–500.*

*12. Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. Eur J Surg. 1995; 161(4):273–81.*

*13. Fenyö G, Lindberg G, Blind P, Enochsson L, Oberg A. Diagnostic decision support in suspected acute appendicitis: validation of a simplified scoring*

*system. Eur J Surg. 1997; 163(11):831–8.*
*14. Karaman K, Ercan M, Demir H, Yalkın Ö, Uzunoğlu Y, Gündoğdu K, et al. The Karaman score: A new diagnostic score for acute appendicitis. Ulus Travma Acil Cerrahi Derg. 2018; 24(6):545–51.*
*15. Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. Arch Surg. 2002; 137(7):799–804; discussion 804.*
*16. Shogilev DJ, Duus N, Odom SR, Shapiro NI. Diagnosing appendicitis: evidence-based review of the diagnostic approach in 2014. West J Emerg Med. 2014; 15(7):859–71.*
*17. Genzor Ríos SJ, Rodríguez Artigas JM, Giménez Maurel T, Vallejo Bernad C, Aguirre Prat N, Miguelena Bobadilla JM. Ultrasonography and the Alvarado score in the diagnosis of acute appendicitis: impact on the negative appendectomy rate. Emergencias. 2016; 28(6):396–9.*
*18. Kamran H, Naveed D, Asad S, Hameed M, Khan U. Evaluation of modified Alvarado score for frequency of negative appendicectomies. J Ayub Med Coll Abbottabad. 2010; 22(4):46–9.*
*19. Sammalkorpi HE, Mentula P, Savolainen H, Leppäniemi A. The Introduction of Adult Appendicitis Score Reduced Negative Appendectomy Rate. Scand J Surg. 2017; 106(3):196–201.*
*20. Yoldas O, Karaca T, Tez M. External validation of Lintula score in Turkish acute appendicitis patients. Int J Surg. 2012;10(1):25–7.*
*21. Chae MS, Hong CK, Ha YR, Chae MK, Kim YS, Shin TY, et al. Can clinical scoring systems improve the diagnostic accuracy in patients with suspected adult appendicitis and equivocal preoperative computed tomography findings? Clin Exp Emerg Med. 2017; 4(4):214–21.*
*22. Tamini N, Santurro L, Chiappetta MF, Gattuso I, Barbieri C, Fattori L, et al. Morbidity after negative appendectomy: a single-centre experience on 627 cases. Eur J Trauma Emerg Surg. 2020; 46(4):859–64.*
*23. Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: a systematic review. BMC Med. 2011;9:139.*
*24. Althoubaity FK. Suspected acute appendicitis in female patients. Trends in diagnosis in emergency department in a University Hospital in Western region of Saudi Arabia. Saudi Med J. 2006; 27(11):1667–73.*
*25. Şenocak R, Kaymak Ş. Diagnostic accuracy of ultrasonography and scoring systems: The effects on the negative appendectomy rate and gender. Ulus Travma Acil Cerrahi Derg. 2020; 26(2):306–13.*

***How to cite this article:***
***Baris Mantoglu, Emre Gonullu, Guner Cakmak, Mertcan Akcay, Merve Yigit, Guven Erdogrul, Kayhan Ozdemir, Fatih Altıntoprak. To what extent are the appendicitis scoring systems effective in preventing negative appendectomy. Comparison of nine scoring systems.*** *Ann Clin Anal Med 2022;13(2):210-214*